图书情报工作
第 62 卷 第 5 期 2018 年 3 月

基于信息熵的新媒体环境下网络节点影响力研究*
——以微信公众号为例

■ 邢云菲[1] 王晰巍[1,2] 韩雪雯[1] 张长亮[1]
[1] 吉林大学管理学院信息管理系 长春 130022 [2] 吉林大学大数据管理研究中心 长春 130022

摘要：[**目的/意义**]对新媒体环境下网络节点影响力进行研究，能够深入剖析信息传播规律，从而有助于采取针对性措施对信息传播进行合理控制。[**方法/过程**]基于信息熵理论构建新媒体环境下网络节点影响力模型，以微信公众号为例进行节点影响力测算，对节点直接影响力、间接影响力和综合影响力进行深入分析，最后运用 Matlab 软件对所构建模型进行仿真分析。[**结果/结论**]新媒体环境下网络节点综合影响力随着连接节点数量和节点间互动频率增加而增大，直接影响力和间接影响力也以不同幅度增长，但当间接影响力信息熵值超过 100 时，直接影响力成为影响节点综合影响力的主要因素。

关键词：信息熵 新媒体 节点影响力 微信
分类号：TP393
DOI：10.13266/j.issn.0252-3116.2018.05.009

1 引言

中国互联网行业最新报告显示，截至 2016 年 12 月，我国手机网民规模达 6.95 亿，增长率连续 3 年超过 10%，手机网民占比高达 95.1%，网民上网设备进一步向移动端集中[1]，移动互联网已成为网络舆情信息传播的主要媒介。随着移动互联网应用的全方位渗透，网民纷纷使用微信、微博、论坛等新媒体工具进行信息传播，针对与其自身利益密切相关的话题发表相关的评论、意愿、态度和表达某种情绪[2]。新媒体环境下如何有效对网络节点进行管理和控制成为信息传播和监管机构关注的新问题。

国内外学者对社会网络中节点影响力展开了一定研究。国外学者 D. Miorandi 等[3]针对社交网络用户影响力问题提出 K-核分析法，并设计一个研究框架分析动态网络中网络节点随时间推移变化的特征及趋势；学者 L. Lü 等[4]提出一种基于 leaderRank 的自由参数的算法量化社交网络用户的影响并与 PageRank 的排名结果进行比较；学者 K. Deanne 等[5]基于博弈论理论构建了一种新的社会网络用户竞争影响模型，通过分析拓扑结构及节点扩散趋势识别意见领袖；学者 V. Arnaboldi 等[6]通过对比分析 Twitter 和 Facebook 上大样本的用户数据，认为具有直接影响关系强度的网络用户在网络信息扩散过程中具有重要作用。国内学者曹学艳等[7]提出了一种完整的网络舆情节点挖掘和分类的技术方法以揭示网络舆情信息传播的结构复杂性、无标度性、子社区结构等特征；学者康伟[8]运用社会网络分析方法研究突发事件舆情传播的网络结构特征，以探究信息的传播路径、传播速度及传播范围影响等问题；学者蒋侃等[9]从信息扩散质量出发构建 WSD-Rank 扩散影响力度量模型，并对关键节点的分布规律及形成内因进行解析；学者王曰芬等[10]设计了舆情数据爬取系统并运用社会网络分析法对网络中关键节点的识别与应用进行研究。纵观现有国内外研究成果，目前国内外学者通过建立模型或采用社会网络分析方法对网络节点影响力进行研究的成果较多，但很少有学者运用信息熵理论分析新媒体环境下网络节点影响力。

本文基于信息熵理论对新媒体环境下网络节点影响力进行识别及测定，通过 Matlab 对新媒体环境下网络节点影响力的大小随节点数量和互动频率变化趋势进

* 本文系国家自然科学基金面上项目“信息生态视角下新媒体信息消费行为机理及服务模式创新研究”（项目编号：71673108）和吉林大学高峰学科（群）建设项目研究成果之一。
作者简介：邢云菲（ORCID：0000-0002-5512-4364），博士研究生，E-mail：787397613@qq.com；王晰巍（ORCID：0000-0002-5850-0126），教授，博士生导师；韩雪雯（ORCID：0000-0001-8477-1969），本科生；张长亮（ORCID：0000-0001-7676-7302），博士研究生。
收稿日期：2017-08-02 修回日期：2017-09-07 本文起止页码：76-86 本文责任编辑：刘远颖

行仿真，本文在研究中试图解决以下3个方面的研究问题：①怎样运用信息熵理论分析新媒体环境下网络节点影响力大小？②如何基于信息熵理论构建新媒体环境下网络节点影响力模型？③如何以微信公众号为例获取数据计算节点影响力并对所构建的模型进行仿真？

本文基于信息熵理论构建新媒体环境下网络节点影响力模型，通过获取微信公众号信息计算网络节点影响力大小，采用仿真分析方法模拟网络节点影响力变化趋势。本文在理论层面为新媒体环境下网络节点影响力的研究提供新的理论研究视角；在实践层面分析新媒体环境下微信公众号网络节点影响力大小并预测节点影响力变化规律，对相关行业机构针对新媒体环境下的信息传播进行有效管控具有重要意义。

2 理论基础

2.1 信息熵

1948年国外学者香农[11]在其著作《通信的数学理论》(*a mathematical theory of communication*)中首次提出信息熵的概念，解决了信息的度量问题。信息熵通过定义信息源的不确定性，针对随机变量取值的不同描述信息含量的多少，其表达式见公式(1)，其中，x_i 表示随机变量，p_i 代表所有输出结果的集合(即概率函数)。

$$H(X) = K[-\log p_i] = -\sum_{i=1}^{n} p_i \log p_i \quad 0 \leqslant p_i \leqslant 1,\ i = (1,2,\cdots,n) \qquad \text{公式(1)}$$

公式(1)所蕴含的主要思想为：假设不确定函数 f 是概率 p 的单调递减函数，则 n 个独立的不确定事件的发生概率应为其和。信息熵理论主要被应用在物理学及统计力学领域，也有部分学者用信息熵理论分析信息规律，少量文献将信息熵应用于计算机科学。国外学者 R. Sangam 等[12]基于信息熵理论讨论了 K 核算法的局限性，并提出一种相似系数来提高聚类精度。学者 Y. Li 等[13]分析搜索引擎上用户评论内容相似性，基于信息熵理论计算用户之间联系权重，并对网络社区进行划分。学者魏志惠等[14]利用信息熵理论检验评价指标体系的合理性，然后利用未确知测度模型对所收集的微博用户数据进行综合评价，最终识别出意见领袖。

2.2 网络节点影响力

网络节点是指在网络中具有信息交换行为的节点，节点可以是工作站、客户、网络用户、个人计算机或服务器[15]。整个网络就是由这些网络节点组成的，把许多的网络节点用通信线路连接起来，形成一定的几何关系。网络节点通过对信息的生产、传递、存储和处理从而达到信息扩散的目的[16]。

网络节点影响力分为直接影响力、间接影响力和综合影响力[17]。其中直接影响力是某一节点 A 对与其具有直接联系的其他节点产生的影响力，把所有与节点 A 直接联系的节点集合具有的影响力汇总，即为该节点 A 的直接影响力；间接影响力为某一节点 A 虽然不与另一节点 B 相连，但是通过其他一个或几个中间节点与节点 B 相连，则节点 A 对节点 B 具有间接影响力；把节点 A 对所有与节点 A 间接联系的节点集合产生的影响力汇总，即为节点 A 的间接影响力。综合影响力为直接影响力和间接影响力的综合评价结果，通过对直接影响力和间接影响力分别设定权重 α、β，且 α＋β＝1 进行运算得出节点 A 在信息传播过程中的综合影响力值，将其作为衡量网络节点影响力的最终指标。

2.3 基于信息熵的新媒体环境下网络节点影响力

新媒体是包括社交媒体、数字媒体、移动媒体、智能通信、网络电视等新技术支撑下的新的数字媒体形态，包括微信、微博、qq、博客、贴吧等，是网络公众发表意见并参与社会活动、进行交流与信息共享的新平台，具有数字化、网络化、多元化、实时性和交互性等多种特征[18]。

新媒体环境下网络节点影响力，是指在新媒体环境下的网络节点在信息传播和信息交换过程中所做出的贡献大小。新媒体环境下网络节点影响力研究可帮助相关部门维护网络秩序，促进相关管理机构对网络平台上的信息传播和信息交互进行更好的引导、监测、预警和控制[19]。

新媒体在信息传播过程中，由于网络节点自身属性特征和其交流渠道的差异，这些网络节点的影响力在信息传播过程中具有很大差异[20]。信息熵是离散随机事件的出现概率，由于信息熵的概念比较模糊，本文将信息熵理解成某种特定信息的出现概率。在新媒体环境下，网络节点的信息熵值越大，说明该节点与其他节点发生联系的概率越大；节点之间的信息熵值越大，说明节点之间联系越紧密[21]。

3 新媒体环境下网络节点影响力模型

3.1 影响力分析框架

本文基于信息熵理论构建节点影响力模型，获取一段时间内新媒体环境下网络节点相关的信息传播样本，通过设计代码进行程序编译计算网络节点影响力

大小，最后运用 Matlab 软件对网络节点影响力变化进行仿真分析，具体步骤如下：①构建模型。运用信息熵理论确定新媒体环境下网络节点直接影响力、间接影响力和综合影响力数学表达，构建新媒体环境下网络节点影响力模型。②构建网络。设定搜索时间为2017年3月14－24日，通过编程对数据源进行样本数据采集，使用Excel和Access进行数据处理，取得联系较紧密的网络节点，并绘制节点网络。③程序编译。通过设计Java编程语言在Java平台输入构建的网络中各节点信息和边的权重信息，依据模型对新媒体环境下网络中节点直接影响力、间接影响力和综合影响力的信息熵值进行计算，进而得出各节点的影响力大小。④仿真分析。以模型为依据，运用Matlab软件对网络节点影响力模型进行仿真分析，绘制节点影响力大小随节点数量和节点间互动频率变化的云图进行可视化操作，分析信息传播过程中节点影响力变化规律。本文构建的分析框架如图1所示：

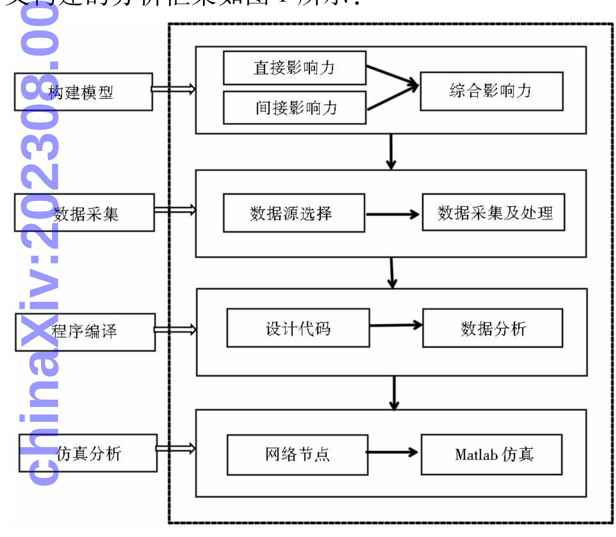


图 1　节点影响力分析框架

3.2　基于信息熵的节点影响力模型构建

3.2.1　直接影响力　在计算节点直接影响力时首先应计算连接节点数量信息熵。$f_{ij}(t)$表示节点 i 与节点 j 在时间 t 的关系。如果节点 i 与节点 j 相连，$f_{ij}(t)=1$，否则$f_{ij}(t)=0$。$K_i(t)$表示在t时刻，所有与节点 i 相连的节点集合，因此 $K_i(t)$是一个衡量在 t 时刻节点 i 的直接影响力的重要指标[22]，即在时间 t 内如果新京报、网易新闻、当时我就震惊了都来源于新华网，则新华网的 $K_i(t)=3$。因此节点 i 的连接节点数 $K_i(t)$表达式为：

$$K_i(t)=\sum_{j=1}^{n}f_{ij}(t) \qquad \text{公式(2)}$$

节点 i 的连接节点信息熵 $I_i^c(t)$表示为：

$$I_i^c(t)=-\sum_{i=1}^{K_i(t)}\frac{1}{K_i(t)+1}\log_{10}\frac{1}{K_i(t)+1} \qquad \text{公式(3)}$$

接下来计算节点间的互动频率信息熵。节点间的互动频率用 $F_{ij}(t)$来表示，代表节点 i 与节点 j 在时间 t 的联系次数，即在时间 t 内如果新华网的文章有两次来源于新京报，则 $F_{ij}(t)=2$。因此节点 i 与节点 j 在时间 t 的互动频率信息熵 $I_i^f(t)$表示为：

$$I_t^f(t)=\sum_{j=1}^{K_i(t)}\frac{F_{ij}(t)}{\sum_{k=1}^{K_i(t)}F_{ik}(t)+1}\log_{10}\frac{F_{ij}(t)}{\sum_{k=1}^{K_i(t)}F_{ik}(t)+1} \qquad \text{公式(4)}$$

节点 i 的直接影响力信息熵 $MI_i(t)$应为其连接的节点数信息熵 $I_i^c(t)$与每个与其相连节点之间的互动频率信息熵 $I_i^f(t)$的乘积，即：

$$MI_i(t)=I_i^c(t)\times I_i^f(t) \qquad \text{公式(5)}$$

3.2.2　间接影响力　在计算节点间接影响力时，首先考虑节点 i 与节点 k 是否通过其他节点相连。$K_{ik}(t)$表示节点 i 与节点 k 都与之相连的节点，即 $K_{ik}(t)=K_i(t)\cap K_k(t)$。若没有其他节点连接节点 i 与节点 k，则 $K_{ik}(t)=0$；若有且只有一个节点 j 连接节点 i 与节点 k，则：$NI_{ik}(t)=NI_{ik}(t)=(\theta+I_{ij})*[1/(\theta+I_{jk})]=(\theta+MI_i(t))*[1/(\theta+MI_j(t))]$。θ为网络节点所处环境影响力，即假设该节点并未与其他节点连接，在计算间接影响力时具有其固有影响力。本文设定 θ＝1。连接节点 i 与节点 k 的路径可能通过一个节点或几个节点，因此计算节点间的间接影响力时考虑单节点连接或多节点连接两种情况。

（1）单节点连接。若有 n 个节点 $j_1,j_2,\cdots j_n$ 连接节点 i 与节点 k，且节点 i 与节点 k 只需要通过一个节点就可以相连，如图2所示，则节点 i 对节点 k 的间接影响力 $NI_{ik}(t)$表示为：

$$NI_{ik}(t)=[\theta+I_{ij}(t)]*[1/(\theta+I_{jk}(t))]+[\theta+I_{in}(t)]*[1/(\theta+I_{mk}(t))]+[\theta+I_{in}(t)]*[1/(\theta+I_{nk}(t))]+[\theta+I_{ip}(t)]*[1/(\theta+I_{pk}(t))] \qquad \text{公式(6)}$$

$$=[\theta+MI_i(t)]*[1/(\theta+MI_j(t))]+[\theta+MI_i(t)]*[1/(\theta+MI_n(t))]+[\theta+MI_i(t)]*[1/(\theta+MI_n(t))]+[\theta+MI_i(t)]*[1/(\theta+MI_p)(t)] \qquad \text{公式(7)}$$

$$=\sum_{j=1}^{N}[\theta+MI_i(t)]*[1/(\theta+MI_j(t))] \qquad \text{公式(8)}$$

（2）多节点连接。若节点 i 与节点 k 不能只通过一个节点连接，需要多个节点才能使信息从节点 i 传递到节点 k，如图3所示。信息从节点 i 传递到节点 k 可以通过(j_1,j_2)或(m_1,m_2,m_3)或(n_1,n_2)，则节点 i 对节点 k 的间接影响力 $NI_{ik}(t)$表示为：

$NI_{ik}(t)=[\theta+I_{ij_1}(t)]*[1/(\theta+I_{j_1j_2}(t))]*[1/(\theta+I_{j_2k}$

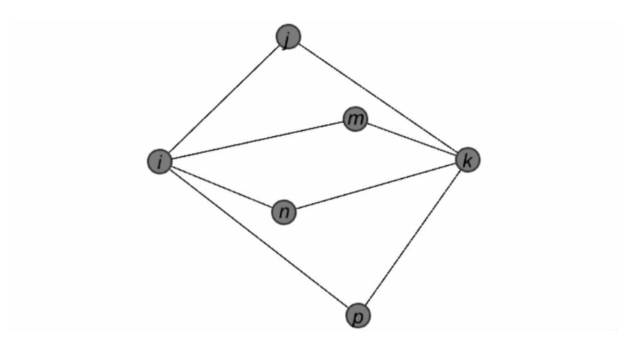


图 2　单节点连接图

$(t))]+[\theta+I_{im_1}(t)]*[1/(\theta+I_{m_1m_2}(t))]*[1/(\theta I_{m_2m_3}(t))]*[1/(\theta+I_{m_2k}(t))]+[\theta+I_{in_1}(t)]*[1/(\theta+I_{n_1n_2}(t))]*[1/(\theta+I_{n_2k}(t))]$　　公式(9)

$=[\theta+MI_i(t)]*[1/(\theta+MI_{j_1}(t)]*[1/(\theta+MI_{j_2}(t))]+[\theta+MI_i(t)]*[1/(\theta+MI_{m_1}(t))]*[1/(\theta+MI_{m_2}(t))]*[1/(\theta+MI_{m_3}(t))]+[\theta+MI_i(t)]*[1/(\theta+MI_{n_1}(t))]*[1/(\theta+MI_{n_2}(t))]$　　公式(10)

$=\sum_{j=1}^{N}[\theta+MI_i(t)]*\{\prod_{j=1}^{n}[1/(\theta+MI_j(t))]\}$　　公式(11)

图 3　多节点连接图

3.2.3　综合影响力　节点的综合影响力应为其直接影响力和间接影响力的综合评价结果，代表某一节点在信息传播过程中产生的作用。如节点 i 只与节点 j 相连，但节点 j 连接成百上千节点，则虽然节点 i 的直接影响力较小，但间接影响力较大，其综合影响力并不小。基于此，节点的综合影响力 $I_i(t)$ 表示为：

$I_i(t)=\alpha MI_i(t)+\beta NI_i(t)$　　公式(12)

其中直接影响力和间接影响力的权重 α 和 β 应满足 α + β = 1。考虑直接影响力和间接影响力的整体评价能力，本文设定权重 α = 0.6，β = 0.4。

3.3　基于微信的网络节点影响力构建

iiMedia Research 数据显示[23]，2016 年中国各大自媒体平台中，微信公众号以 63.4% 的绝对优势领衔自媒体行业，微博自媒体平台成为传播第二渠道选择，其占比为 19.3%。艾媒咨询分析师认为，微信公众号作为关注才能阅读的相对封闭化的自媒体，在一定程度上激发了今日头条、UC 等新媒体的崛起。微信公众号作为信息传播的一大途径，也成为传统新闻媒体转型的发展领域。微信作为新兴的新媒体社交平台，具有广泛的信息传播渠道和多样的附加功能[24]。微信公众号自 2011 年 1 月推出以来发展迅速。截至 2016 年 8 月，微信公众账号已经超过了 1 400 万个。

微信公众号对新媒体网络提供业务服务，主要面向企业、媒体、政府等。微信公众号平台凭借信息推送、评论、相关信息链接等功能迅速成为新媒体环境下网民分享交流的重要渠道，形成了集多种信息交流方式于一体的信息交流平台[25]。由于微信公众号平台上的文章转发量、评论及路径都仅为运营者可见，出于对微信（如自媒体或微商）隐私保护的考虑，其相关数据不完全公开。基于此，本文首先通过搜狗微信搜索引擎以“网易新闻”为关键字获取微信公众号及其文章地址，再借助“微指数”平台并使用爬虫工具进行数据采集及获取，最后采集到 38 个公众号在 2017 年 3 月 21－24 日发表的 245 篇文章，选择其中 7 个联系较紧密的公众号构成全联通网络，即每个节点既是“来源”节点又是“被来源”节点，然后使用 Access、Excel 工具进行数据处理，根据其节点、文章来源关系构建的初始网络，如图 4 所示：

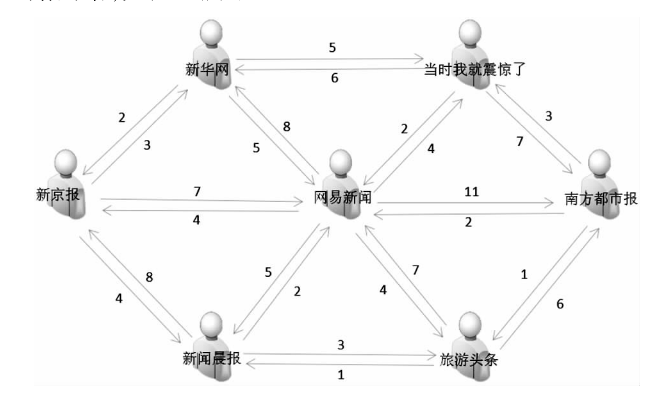


图 4　节点影响力初始网络图

公众号样本分别为新华网、当时我就震惊了、新京报、网易新闻、南方都市报、新闻晨报和旅游头条，有向边代表微信公众号文章来源方向，数字代表文章来源次数，如新华网有 5 篇文章来源于当时我就震惊了，有 2 篇文章来源于新京报，即节点间互动频率。

4　数据结果及讨论分析

4.1　节点影响力测算结果

4.1.1　直接影响力　在计算新媒体环境下网络节点直接影响力时，通过 Java 编程计算出本文构建的微

信公众号网络中各节点的连接节点数量信息熵和节点间互动频率信息熵，其直接影响力值为这两项的乘积。经计算，直接影响力大小由低到高的节点分别是网易新闻、新京报、新华网、南方都市报、当时我就震惊了、新闻晨报和旅游头条。其中网易新闻的直接影响力最大，为 0.660 37，因为与其直接连接的节点有 6 个，其他节点的直接连接节点数均为 3，因此网易新闻的连接节点信息熵和与其他节点间互动频率信息熵值都较大，分别为 0.903 089 和 0.731 234，远高于其他 6 个节点，因此网易新闻的直接影响力最大。其余 6 个节点的直接影响力值比较相近，因为每个节点的直接连接节点数相同，其连接节点信息熵值均为0.421 758。由于这 6 个节点的互动频率不同，因此互动频率信息熵值不同，直接影响力的值具有一定差异，但数值比较接近，都在 0.16 – 0.19 之间。旅游头条的直接影响力最小，为 0.165 348。如表 1 所示：

表 1 直接影响力统计

排序	网络节点	连接节点数量信息熵	互动频率信息熵	直接影响力
1	网易新闻	0.903 089	0.731 234	0.660 37
2	新京报	0.421 758	0.447 333	0.188 666
3	新华网	0.421 758	0.444 274	0.187 376
4	南方都市报	0.421 758	0.433 881	0.182 993
5	当时我就震惊了	0.421 758	0.429 697	0.181 228
6	新闻晨报	0.421 758	0.402 966	0.169 954
7	旅游头条	0.421 758	0.392 046	0.165 348

4.1.2 间接影响力 在本文构建的微信公众号网络中，有 24 条连接任意两节点的线段。除了网易新闻的直接连接节点有 6 个，其他节点的直接连接节点均为 3 个。经计算相邻两个节点间的互动频率信息熵值见表 2，当时我就震惊了发布的文章来源于新华网的信息熵值最大，为 0.159 176，旅游头条发布的文章来源于新闻晨报的信息熵值最小，为 0.081 866，是唯一一条连接节点信息熵值低于 0.1 的信息传播路径。通过 Java 编程输入各节点连接节点数量和相邻两节点间互动频率信息熵，计算构建的微信公众号网络中 6 个节点的间接影响力，得到统计结果见表 3。网易新闻的间接影响力值最高，为0.474 152。其他公众号节点间接影响力大小按降序排列分别为当时我就震惊了，新京报、新华网、新闻晨报、南方都市报和旅游头条。上述 6 个公众号的间接影响力值均在 0.23 – 0.27 之间。

表 2 相邻两个节点间互动频率信息熵

源节点→目标节点	信息熵	源节点→目标节点	信息熵
新华网→新京报	0.129 692	南方都市报→旅游头条	0.129 692
新华网→网易新闻	0.158 421	南方都市报→网易新闻	0.159 040
新华网→当时我就震惊了	0.158 421	南方都市报→当时我就震惊了	0.150 515
新京报→新华网	0.143 359	当时我就震惊了→南方都市报	0.154 464
新京报→网易新闻	0.150 515	当时我就震惊了→网易新闻	0.116 675
新京报→新闻晨报	0.155 448	当时我就震惊了→新华网	0.159 176
新闻晨报→新京报	0.129 756	网易新闻→新华网	0.145 158
新闻晨报→网易新闻	0.125 064	网易新闻→新京报	0.106 027
新闻晨报→旅游头条	0.146 959	网易新闻→新闻晨报	0.119 074
旅游头条→新闻晨报	0.081 866	网易新闻→旅游头条	0.106 027
旅游头条→网易新闻	0.150 515	网易新闻→南方都市报	0.157 334
旅游头条→南方都市报	0.157 704	网易新闻→当时我就震惊了	0.106 027

表 3 间接影响力统计

排序	网络节点	间接影响力
1	网易新闻	0.474 152
2	当时我就震惊了	0.269 784
3	新京报	0.258 348
4	新华网	0.246 714
5	新闻晨报	0.245 419
6	南方都市报	0.237 926
7	旅游头条	0.230 238

4.1.3 综合影响力 本文对新媒体环境下网络节点综合影响力计算结果见表 4（直接影响力权重系数 α 取值0.6，间接影响力权重系数 β 取值0.4）。综合影响力从大到小排列分别为网易新闻、新京报、新华网、当时我就震惊了、南方都市报、新闻晨报和旅游头条。其中，网易新闻的综合影响力达到 0.585 883，远高于其他 6 个节点。因此，在本文构建的微信公众号网络中，网易新闻是核心节点。其他 6 个节点的综合影响力值均在 0.19 – 0.22 之间。本文基于微信公众号来源数据构建的网络具有延展性，当节点数量更多，节点之间联系更复杂时，运用本文建立的网络节点影响力测算模型可以处理更庞大数据，对网络舆情信息传播过程中核心节点分类和意见领袖识别具有重要作用。

4.2 仿真分析

Matlab 是被用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境，对于矩阵运算、绘制函数和数据、实现算法具有较强的处理能力。本文运用 Matlab 软件对新媒体环境下网络节

表 4 综合影响力统计

排序	网络节点	综合影响力
1	网易新闻	0.585 883
2	新京报	0.216 539
3	新华网	0.211 111
4	当时我就震惊了	0.216 65
5	南方都市报	0.204 966
6	新闻晨报	0.200 14
7	旅游头条	0.191 304

点的直接影响力、间接影响力和综合影响力变化趋势绘制云图进行可视化分析。

4.2.1 直接影响力仿真

（1）连接节点数量信息熵。在 Matlab 中把新媒体环境下网络节点影响力信息熵的数学表达式转化为 C 语言进行编程，用 scatter(x,y)绘制连接节点数量信息熵(y 轴)随自变量节点数量(x 轴)增加的变化趋势散点图得到仿真结果如图 5 所示。由图可知，连接节点数量信息熵随某网络节点连接的其他节点数量增多而增大，但增速随网络规模变化呈现明显差异。当连接节点数量小于 700 时，信息熵增速较快，大于 700 时则较慢。因此对连接节点数量小于 700 和大于 700 时的节点连接数量信息熵变化分别拟合趋势线，如图 6(a)、(b)所示。可见，当连接节点数量小于 700 时，其趋势线为标准对数，公式为 $y = 0.4343\ln(x) - 1E-13$ ($R^2 = 1$)；当连接节点数量大于 700 时，其趋势线公式为 $y = 0.000\ 01x + 2.489$ ($R^2 = 0.997$)。当网络中某一原创节点连接节点数量较少时，其连接节点信息熵随节点数量增加的速度较快，增速为 0.434 3；当连接节点数量较多时增速缓慢，小于 0.001，连接节点数量信息熵值趋近于 3。R 方值为 0.997，说明当连接节点数量超过 700 时，其信息熵大小与节点连接数量呈高度线性相关。由此得出，在新媒体环境下信息传播过程中，当与某一节点连接的节点数从 0 开始增加时，其连接节点数量信息熵值呈对数增长趋势；但当与其连接的节点数量超过 700 时，其连接节点数量信息熵值呈线性增长趋势。

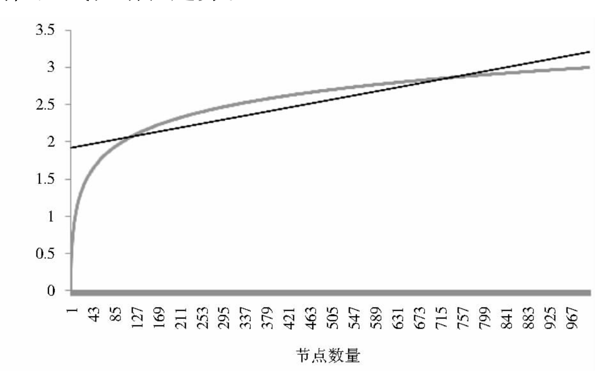
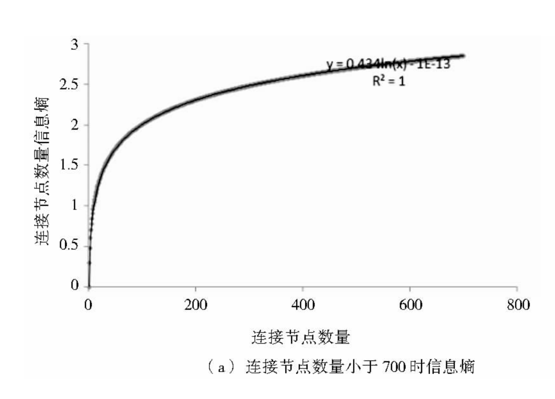
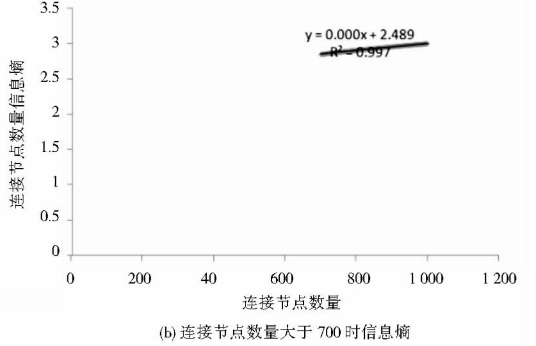


图 5 连接节点数量信息熵



(a) 连接节点数量小于 700 时信息熵



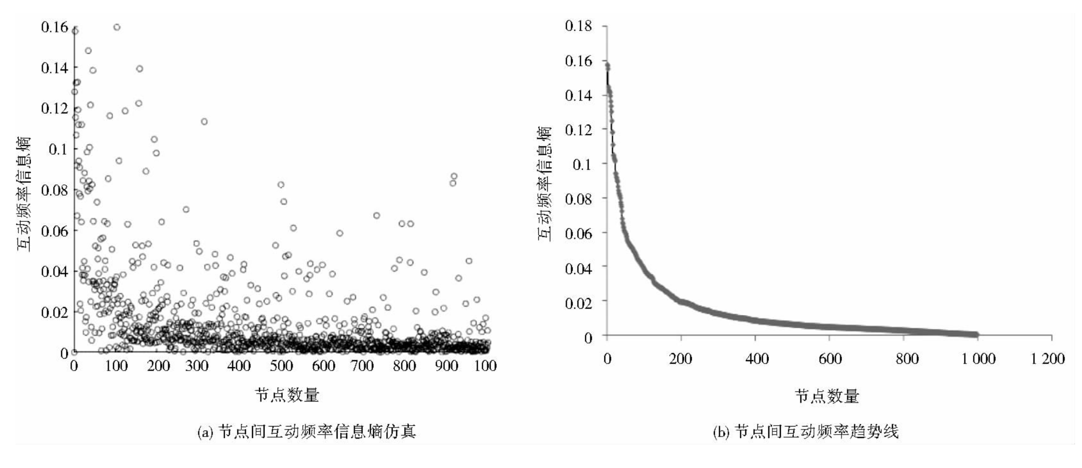
(b) 连接节点数量大于 700 时信息熵

图 6 连接节点数量信息熵变化

（2）互动频率信息熵。通过在 Matlab 拟合新媒体环境下网络节点的节点互动频率信息熵得到仿真图如图 6(a)、(b)所示。由图 7(a)可知，网络中节点互动频率信息熵的值与节点数量呈负相关。网络中节点数量小于 200 时，其互动频率信息熵值在 0－0.16 之间，分布相对分散；当节点数量大于 200 时，其互动频率信息熵值主要集中在 0－0.02 之间，分布相对密集，且随着节点数量增加有更加集中并下降的趋势。筛选排除互动频率信息熵值低于 0.02 的点，其他数值按降序排列，得到一个相对标准的幂律分布曲线，如图 7(b)，其幂律分布多项式公式为 $y = 0.535\ 6x - 0.6$ ($R^2 = 0.895$)。

由此得出，在新媒体环境下信息传播过程中，节点间互动频率信息熵的值随节点数量增多而下降。当节点数量小于 200 时，该网络中节点互动频率信息熵值相对较大，但不超过 0.16；当节点数量大于 200 时，该

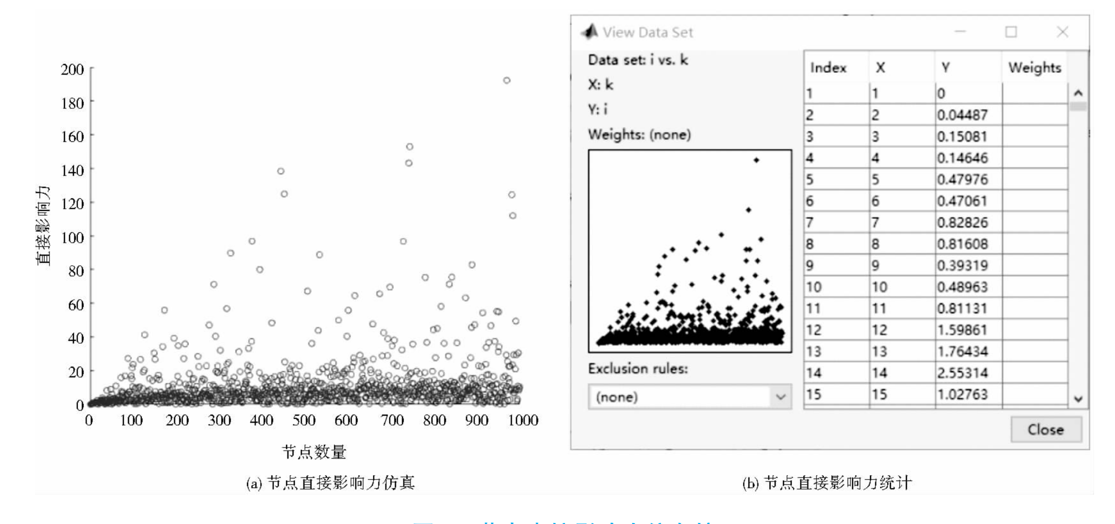
网络中节点互动频率信息熵值趋近于 0。但由于其变化趋势微弱且数值相近,在计算新媒体环境下网络节点直接影响力时,连接节点数量信息熵对其的影响更大。因此,政府及相关舆情管控部门在对网络舆情信息传播监管过程中,相比较增加对某一意见领袖的管控次数,与更多关键节点产生联系并进行舆情监控对政府提高其直接影响力具有更大作用。



(a) 节点间互动频率信息熵仿真　(b) 节点间互动频率趋势线

图 7　节点间互动频率信息熵

(3) 直接影响力仿真。新媒体环境下网络节点的直接影响力为该节点的连接节点数量信息熵与该节点与其他节点的互动频率信息熵的乘积,通过在 Matlab 中拟合新媒体环境下网络节点的直接影响力仿真结果见图 8(a)。由图可知,网络节点的直接影响力与连接节点数量呈正相关,且数值分布逐渐分散。节点直接影响力与其连接数量变化统计表见图 8(b)。原创节点的直接影响力峰值点随连接节点数量增加呈现线性递增趋势,即原创节点的阈值与连接节点数量呈正相关,拟合趋势线公式为 $y = 0.2406x - 1.485$ ($R^2 = 0.873$)。通过在 Matlab 用 C 语言输入节点直接影响力信息熵表达式,用 scatter(x,y,z) 绘制节点直接影响力信息熵(z 轴)随自变量节点数量(x 轴)和节点间互动频率(y 轴)变化的散点图如图 9 所示。网络中节点的直接影响力随着连接节点数量和与其他节点互动频率的增加而增大,但大部分节点的直接影响力信息熵值分布较均匀且数值水平较低。因此在新媒体环境下信息传播过程中,与某一原创节点连接的节点数量越少且互动频率越低,其直接影响力越小;与该原创节点连接的节点数量越多且互动频率越高,其直接影响力越大。在舆情管控过程中,政府及相关部门应同时提高与关键节点的联系频率及管理的节点数量,才能最优提高管控力度。同时,直接影响力的阈值趋势线也说明对网络节点的管控具有一定的限制,超过该阈值范围,对这些节点的管理也会失去效果。



Index	X	Y	Weights
1	1	0	
2	2	0.04487	
3	3	0.15081	
4	4	0.14646	
5	5	0.47976	
6	6	0.47061	
7	7	0.82826	
8	8	0.81608	
9	9	0.39319	
10	10	0.48963	
11	11	0.81131	
12	12	1.59861	
13	13	1.76434	
14	14	2.55314	
15	15	1.02763	

(a) 节点直接影响力仿真　(b) 节点直接影响力统计

图 8　节点直接影响力信息熵

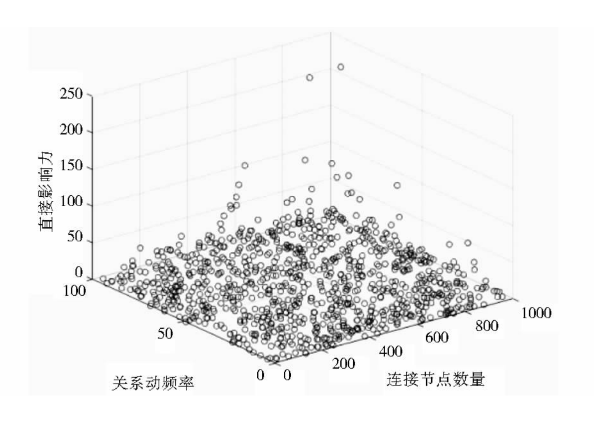
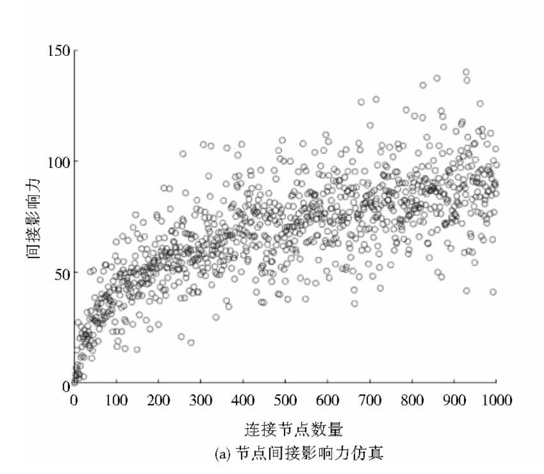


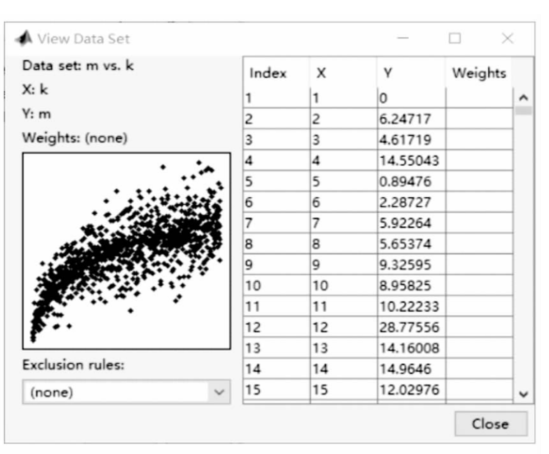
图 9　节点直接影响力仿真

4.2.2　间接影响力仿真　根据新媒体环境下网络节点直接影响力计算结果，在 Matlab 拟合新媒体环境下网络节点的间接影响力变化趋势得到仿真图见图 10(a)。某原创节点间接影响力信息熵值随连接节点数量变化的统计表见图 10(b)。由图可知，当某一原创节点的连接节点数量小于 500 时，间接影响力信息熵值增长速度相对较快，其增长趋势符合对数分布，公式为 $y = 0.570\,2\ln(x) + 6E-12$ ($R^2 = 0.892$)，当原创节点的连接节点数量大于 500 时，随着连接节点数量增多，间接影响力信息熵值增速减慢，趋近于 120。某原创节点的间接影响力随节点数量变化和互动频率变化的 3D 仿真结果见图 11。信息传播过程中，节点的间接影响力随连接节点数量和与其他节点互动频率的增加而增大。在计算新媒体环境下网络节点综合影响力时，当原创节点连接节点数量超过 700 个时，其间接影响力对其综合影响力的影响指数较低。

因此在新媒体环境下信息传播过程中，政府及相关部门可以通过增加管理数量和管理次数对信息的间接传播进行引导或控制，提高管理部门的间接影响力。但当管控的节点数量超过一定范围且管理次数超过一定频率，对提高其间接影响力作用甚微，这时应把管理重心放在提高其直接影响力上。



(a) 节点间接影响力仿真



Index	X	Y	Weights
1	1	0	
2	2	6.24717	
3	3	4.61719	
4	4	14.55043	
5	5	0.89476	
6	6	2.28727	
7	7	5.92264	
8	8	5.65374	
9	9	9.32595	
10	10	8.95825	
11	11	10.22233	
12	12	28.77556	
13	13	14.16008	
14	14	14.9646	
15	15	12.02976	

(b) 节点间接影响力统计

图 10　节点间接影响力信息熵

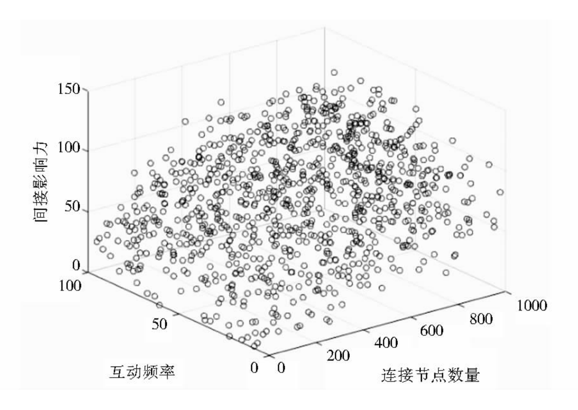
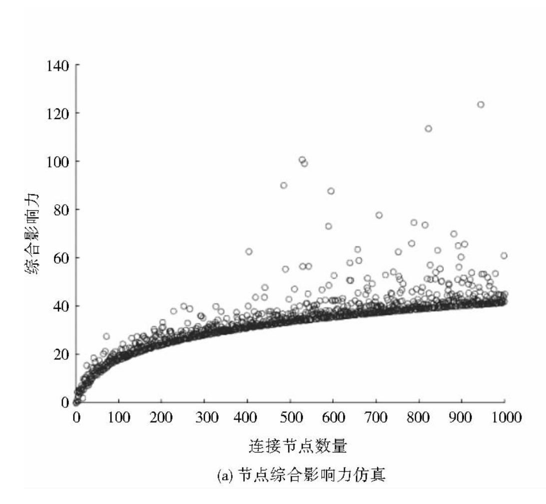
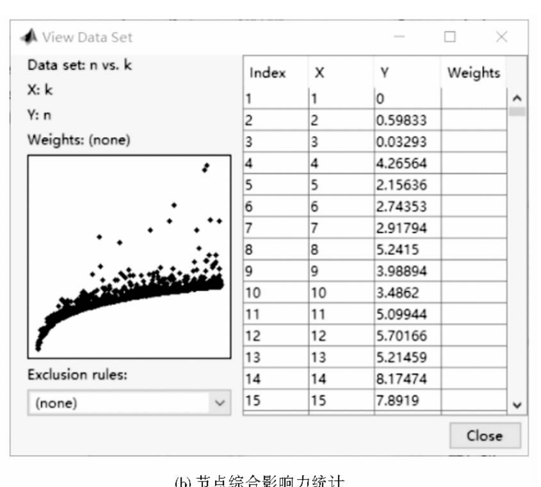


图 11　节点间接影响力仿真

4.2.3　综合影响力仿真　在 Matlab 中对上文计算得到的直接影响力和间接影响力分别设定 $\alpha = 0.6$ 和 $\beta = 0.4$ 的权重绘制新媒体环境下网络节点综合影响力仿真结果见图 12(a)。该图与节点直接影响力仿真图相似，但数值变化幅度不同。某原创节点的综合影响力信息熵值随连接节点数量变化的统计表如图 12(b) 所示。节点的综合影响力与连接节点数量呈正相关，且数值分布逐渐分散。对该仿真图中节点综合影响力数值拟合趋势线，公式为 $y = 9.391x^{0.263\,4} - 12.48$ ($R^2 = 0.721\,8$)。新媒体环境下网络节点综合影响力大小在连接节点数量较少时增速较快，但随着连接节点数量增加，虽然该节点的综合影响力大小还保持增长状态，其增速减慢。



(a) 节点综合影响力仿真

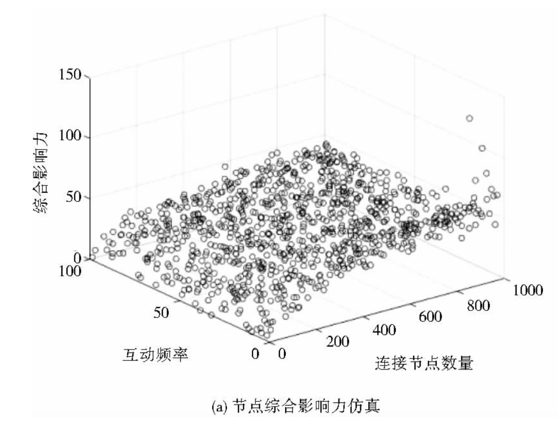


Index	X	Y	Weights
1	1	0	
2	2	0.59833	
3	3	0.03293	
4	4	4.26564	
5	5	2.15636	
6	6	2.74353	
7	7	2.91794	
8	8	5.2415	
9	9	3.98894	
10	10	3.4862	
11	11	5.09944	
12	12	5.70166	
13	13	5.21459	
14	14	8.17474	
15	15	7.8919	

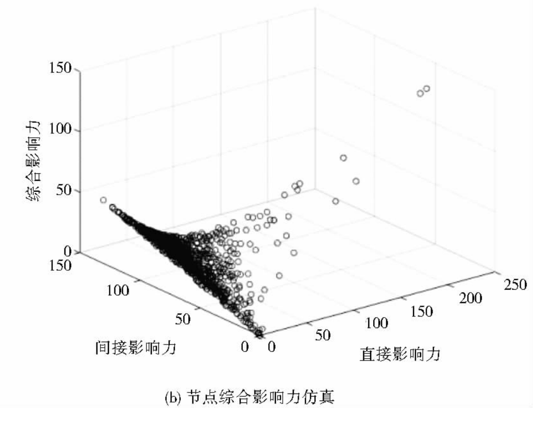
(b) 节点综合影响力统计

图 12　节点总和影响力信息熵

新媒体环境下网络节点综合影响力随连接节点数量和互动频率变化的仿真图见图 13(a)。随着原创节点连接节点数量和互动频率增加,其综合影响力也相对增加,但互动频率变化对节点综合影响力的影响要弱于连接节点数量变化对节点综合影响力的影响。在 Matlab 中把直接影响力和间接影响力分别设定为两个变量,绘制新媒体环境下网络节点的综合影响力随着直接影响力和间接影响力双重变量变化的散点仿真结果见图 13(b)。在直接影响力和间接影响力的共同作用下,综合影响力呈上升趋势。当某一原创节点间接影响力较小时,直接影响力与间接影响力的作用程度相近,但当间接影响力大小超过 100 时对综合影响力结果产生微弱影响,直接影响力成为影响新媒体环境下网络节点综合影响力的主要因素。因此,当某个网络中节点数量和节点间互动频率超过一定范围时,舆情管理者在评价网络节点影响力过程中应以直接影响力结果为主对节点影响力进行识别和分类。



(a) 节点综合影响力仿真



(b) 节点综合影响力仿真

图 13　节点综合影响力信息熵

5　研究结论

本文在理论层面,基于信息熵理论构建新媒体环境下网络节点影响力模型,确定了连接节点数量信息熵、互动频率信息熵,建立了直接影响力信息熵、间接影响力信息熵、综合影响力信息熵的数学表达。通过 Matlab 软件对模型进行仿真,并针对仿真结果展开系统分析。数据结果表明,随着新媒体环境下网络节点的连接节点数量和互动频率增加,其直接影响力和间接影响力大小都随之增加;但增加的幅度不同,直接影响力增加幅度高于间接影响力。综合影响力大小也随着连接节点数量和互动频率增加而增加,且当节点数量较少时综合影响力增加速度较快,当节点数量持续增多并超过一定范围,综合影响力增加速度减慢;当间接影响力信息熵值达到 100 时将不对综合影响力产生重要作用,直接影响力成为影响其综合影响力的主要因素。本文的研究为新媒体环境下网络节点影响力研究提供新的理论基础和模型分析方法。

在实践层面,本文以微信公众号为例,绘制云图预测节点数量增多时影响力大小的变化趋势,对政府及相关部门更好地监管新媒体环境下网络信息传播提供有力支持。政府及相关管理者可基于本文提出的网络节点影响力模型对网络中节点影响力大小进行测定,实时监控网络中节点数量和节点间互动频率的变化,并重点管理影响力较大的节点,从而对信息传播进行有效监控和管理,防止新媒体环境下不良信息扩散。

本文在研究过程中也存在一定的局限性,仅选择7个微信公众号的来源数据对新媒体环境下网络节点的影响力进行测算,其样本数据覆盖不够广泛。在后续研究中,本文将选择大样本数据,如微博转发评论信息对该模型进行验证,并对新媒体环境下信息传播进行预警分析,进而对比分析不同新媒体环境下网络节点信息传播影响力的异同。此外,还要继续挖掘信息熵理论在信息传播等其他方面的应用。

参考文献:

[1] 中国互联网络信息中心. 第39次《中国互联网络发展状况统计报告》[EB/OL]. [2017－01－22]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201701/t20170122_66437.htm.
[2] KIM K, BAEK Y M, KIM N. Online news diffusion dynamics and public opinion formation: a case study of the controversy over judges' personal opinion expression on SNS in Korea[J]. The social science journal, 2015,52(2):205－216.
[3] Miorandi D, Pellegrini F D. K-shell decomposition for dynamic complex networks[C]// Proceedings of the international symposium on modeling and optimization in mobile, ad hoc and wireless networks. Paris: IEEE, 2010:488－496.
[4] LYU L, ZHANG Y, YEUNG C H. Leaders in social networks, the delicious case[J]. PLos ONE, 2011,6(6):e21202.
[5] DEANNE K, KATHARINE H, HONERT R, et al. Nuclear power in Australia: a comparative analysis of public opinion regarding climate change and the Fukushima disaster[J]. Energy policy, 2014, 65:644－653.
[6] ARNABOLDI V, CONTI M, LA GALA M. Ego network structure in online social networks and its impact on information diffusion [J]. Computer communications, 2016,76:26－41.
[7] 曹学艳,段飞飞,方宽,等. 网络论坛视角下突发事件舆情的关键节点识别及分类研究[J]. 图书情报工作,2014,58(4):65－70.
[8] 康伟. 突发事件舆情传播的社会网络结构测度与分析——基于"11·16校车事故"的实证研究[J]. 中国软科学,2012(7):169－178.
[9] 蒋侃,唐竹发,隋浩. 基于微博信息扩散质量的网络舆情关键节点识别[J]. 情报科学,2016,34(7):64－69.
[10] 王曰芬,杭伟,梁丁洁. 微博舆情社会网络关键节点识别与应用研究[J]. 情报资料工作,2016(3): 6－11.
[11] SHANNON C E. A mathematical theory of communication[J]. Bell system technical journal, 1948,27: 379－423. doi:10.1002/j.1538－7305.1948.tb01338.x.
[12] SANGAM R, OM H. The k-modes algorithm with entropy based similarity coefficient[J]. Computer science, 2015,50:93－98.
[13] LI Y, SHA Y, SHAN J. The research of weighted community partition based on SimHash [J]. Computer science, 2013, 17: 797－802.
[14] 魏志惠,何跃. 基于信息熵和未确知测度模型的微博意见领袖识别——以"甘肃庆阳校车突发事件"为例[J]. 情报科学,2014,32(10):38－43.
[15] 陈远,李韫慧,张敏. 基于节点度测度SNS用户信息传播贡献的实证研究——以腾讯微博为例[J]. 情报杂志,2014,33(10):159－164.
[16] MANKU G S, JAIN A. Detecting near-duplicates for Web crawling [M]. New York: ACM, 2007.
[17] CHARIKAR M S. Similarity estimation techniques from rounding algorithms[M]. New York: ACM, 2002.
[18] 武澎,王恒山. 基于特征向量中心性的社交信息超网络中重要节点的评判[J]. 情报理论与实践,2014,37(5):107－113.
[19] 王晰巍,邢云菲,赵丹,等. 移动环境下网络舆情信息传播路径及传播规律研究[J]. 情报理论与实践,2016,39(9):107－113.
[20] KIM K, BAEK Y M, KIM N. Online news diffusion dynamics and public opinion formation: a case study of the controversy over judges' personal opinion expression on SNS in Korea[J]. The social science journal, 2015,52(2):205－216.
[21] 王晰巍,邢云菲,赵丹,等. 基于社会网络分析的移动环境下网络舆情信息传播研究——以新浪微博"雾霾"话题为例[J]. 图书情报工作,2015,59(7):14－22.
[22] PENG S, YANG A, Cao L. Social influence modeling using information theory in mobile social networks [J]. Information science, 2017,379:146－159.
[23] 艾媒报告. 2017年中国新媒体行业全景报告[EB/OL]. [2017－03－01]. http://www.iimedia.cn/50347.html.
[24] 中国社会科学网. 中国新媒体发展报告 [EB/OL]. [2017－09－01]. http://www.cssn.cn/zx/bwyc/201706/t20170626_3560419_1.shtml.
[25] 方婧,陆伟. 微信公众号信息传播热度的影响因素实证研究[J]. 情报杂志,2016,35(2):157－162.

作者贡献说明:

邢云菲:论文撰写及修改;
王晰巍:提出研究命题、研究思路,修订并确定论文最后版本;
韩雪雯:数据采集及处理;
张长亮:英文文献收集及摘要翻译。

Research on the Influence of Network Nodes in New Media Environment Based on Information Entropy: A Case Study of WeChat

Xing Yunfei[1] Wang Xiwei[1,2] Zhang liu[1] Li Shimeng[1]
[1] School of Management, Jilin University, Changchun 130022
[2] Big Data Management Research Center, Jilin University , Changchun 130022

Abstract: [**Purpose/significance**] The research on the influence of network nodes in new media environment can deeply analyze the law of information dissemination, and thus help to take appropriate measures to control the information dissemination reasonably. [**Method/process**] This paper constructs the model of users' node influence in the new media environment based on information entropy theory, taking WeChat as an example to analyze the influence of network nodes. By using java programming method, the paper aims to calculate the numerical value of network nodes, including direct influence, indirect influence and total influence. And finally it use Matlab software to simulate the model. [**Result/conclusion**] Network nodes' total influence increases with the growth in the number of connections and the frequency of interactions between nodes. Direct influence and indirect influence are growing at a different rate, but the direct influence becomes the main factor affecting the comprehensive effect of nodes when the value of indirect influence is over 100.

Keywords: information entropy new media node influence Wechat